Can you pray your way towards statistical significance? An experimental test

Department of Math & Theology (Matheology)[1]

1 - PNIS

Introduction

It might surprise you to learn that the effect of prayer in determining particular outcomes—like the recovery of a patient—is a relatively popular research topic. A Google Scholar search for "effect of prayer" (without quotes) returns 357,000 results. For comparison, "effect of selenium" (selenium being a necessary component of human life, as well as a potentially toxic element) yields 762,000 results, and "effect of beryllium" (beryllium being a particularly harmful heavy metal) yields only 163,000. Considering that the majority of prayer studies use human subjects and that the majority of heavy metal studies use non-human subjects, we might know more about how humans are affected by prayer than by many potentially harmful substances.

The numerous studies on the efficacy of prayer give conflicting results. In support of prayer, Byrd (1988) found that patients that were prayed for had better outcomes than a control group. Byrd's study was successfully replicated by Harris et al. (1999), with a somewhat improved methodology. On the contrary, results from the STEP project (Study of the Therapuetic Effects of Intercessory Prayer), published in 2006, showed that prayer did not significantly improve patient outcomes, and may actually have had negative effects. And, of course, in the middle are the various meta-analyses of prayer studies, which have provided equivocal results (Astin et al. 2000, Powell et al. 2003, Masters et al. 2006, Hodge 2007).

While most prayer studies use patient health as the prayed-for outcome, no study to our knowledge has examined if prayer can affect arguably the most important outcome for scientists themselves—achieving statistical significance (Fanelli 2012, Economist). Statistical significance (at the alpha level of 0.05, of course) is so important to scientists that it is seen as an object of worship (Neuhauser and Provost 2012), and even approaches cult status (Ziliak and McCloskey 2008). Judging by the miraculous preponderance of studies with p-values that are juuust significant (reviewed by Masicampo and Lalande 2012, Francis et al. 2014, and de Winter and Dodou 2015; also see here), such worshipping appears to be paying off.

Thus, an important question for scientists is whether praying can actually help you achieve statistical significance. The ramifications of this power are hard to overstate. If researchers can successfully pray for statistical significance, then the very idea of proper experimental design (such as adequate sample size or proper controls), and maybe even the very idea of the Scientific Method—the backbone of Science—would be rendered moot. What would be the point of conducting a study to test any hypothesis when you can just pray for significance? Besides, all scientists pray before running their statistics, so why not test to see if it actually works?

Thus, to save science, we test whether or not saying a prayer can help you achieve statistical significance before running a statistical test.

Methods

General Design – We tested two distinct times in which a researcher might be inclined to pray for statistical significance. The first opportunity occurs before any data is collected, with a prayer asking God to deliver a dataset that contains some statistically significant finding. Here, the divine intervention lies in helping obtain a significant dataset; the onus is on the researcher to actually find what is significant. We call this the "A Priori" scenario.

The second opportunity occurs after all data has been collected without *a priori* prayer, and the researcher is just about to conduct a statistical test. Here, the divine intervention is presumably stronger—God is asked to make a possibly non-significant dataset significant. We call this the "Desperation" scenario. It should also be noted that the Desperation Scenario is likely to be more applicable to researchers.

Treatments – We employed one treatment (prayer) and two controls. Prayers were obtained from two websites: here for the A Priori Scenario and here for the Desperation Scenario. For each prayer, we added a phrase asking for God's assistance in helping find statistical significance. To control for the act of saying words, we generated random text from an online Random Text Generator, with the text having the same number of words as each prayer. Lastly, we had a control where nothing was said before conducting the statistical test.

Experiments – Our experimental design was built around the generation of random datasets.

For the A Priori Scenario, we followed these steps:

    1. Apply treatment (prayer, random text, or nothing).
    2. Generate two random datasets.
    3. Use Student's *t*-test to determine if the two datasets had significantly different means.

For the Desperation Scenario, we followed these steps:

    1. Generate 200 random datasets.
    2. Randomly choose two datasets.
    3. Apply treatment (prayer, random text, or nothing).
    4. Use Student's *t*-test to determine if the two datasets had significantly different means.

Prayers and random text were said aloud by one of the authors, who was raised Catholic.

For both scenarios, the datasets were specifically designed so that any two datasets had a 50% chance of being significantly different at an alpha level of 0.05 (this was tested beforehand using simulations). To reflect the small sample sizes popular among scientists, each dataset contained 50 observations. To further reflect modern scientific experiments, we replicated each treatment only 10 times.

Statistical Analysis – After each Student's *t*-test, we recorded whether or not the two datasets differed significantly (i.e., if $P < 0.05$), and the effect size (using Cohen's *d*). For our count response variable (significant/not significant) we used a 2 x 3 contingency table and χ^2 test. For our continuous response variable (value of Cohen's *d*), we used a one-way ANOVA. Also, we did not ourselves pray before applying these statistical tests.

Results

A Priori Scenario – Descriptive statistics for each treatment are given in Table 1. Our contingency table was significant ($\chi^2 = 9.6$, df $= 2$, $P = 0.008$). Qualitatively, the Prayer and Nothing treatments resulted in more statistically significant outcomes than the Random Text treatment. Our one-way ANOVA was also sig-

Table 1. Descriptive statistics for each treatment in the A Priori Scenario (prayer before data collected).

Statistic	Prayer	Random Text	None
# Trials	10	10	10
# Sig. Results	7	1	7
mean Cohen's *d*	0.483	0.219	0.488
var. Cohen's *d*	0.027	0.021	0.017
mean *P*-value	0.051	0.366	0.033
var. *P*-value	0.004	0.063	0.001

Table 2. Descriptive statistics for each treatment in the Desperation Scenario (prayer after data collected).

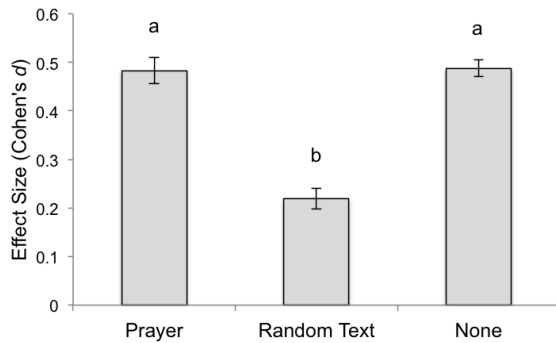| Statistic	Prayer	Random Text	None
# Trials	10	10	10
# Sig. Results	2	5	5
mean Cohen's *d*	0.290	0.420	0.343
var. Cohen's *d*	0.022	0.032	0.046
mean *P*-value	0.241	0.098	0.255
var. *P*-value	0.057	0.013	0.133

Figure 1. Cohen's *d* for the three treatments in the A Priori Scenario. Treatments labeled with different letters indicate a significant difference.

Figure 2. Cohen's *d* for the three treatments in the Desperation Scenario. Treatments labeled with different letters indicate a significant difference.

nificant ($F_{2,27}$ = 12.36, P < 0.001). A Tukey's HSD test showed that Cohen's *d* was significantly higher in the Prayer and Nothing treatment than the Random Text treatment (Fig. 1).

Desperation Scenario – Descriptive statistics for each treatment are given in Table 2. Our contingency table was not significant (χ^2 = 2.5, df = 2, P = 0.2865). Our one-way ANOVA was also not significant ($F_{2,27}$ = 1.29, P = 0.293; Fig. 2). Nothing to see here.

Discussion

Surprise, surprise, our definitive study on the effects of prayer in helping attain statistical significance gave mixed results. On one hand, praying before generating a dataset resulted in more significant differences than reciting random text. On the other hand, praying did not perform better than simply doing nothing. Plus, praying had no effect on statistical significance after the data had already been collected (i.e., the Desperation Scenario).

For atheist scientists, the implications are pretty straightforward. You can probably continue doing what you've always been doing. You're probably not praying at any point during your research projects and we can conclude from our experiment that God is okay with that. Just don't recite random text; it seems God does not like to be mocked.
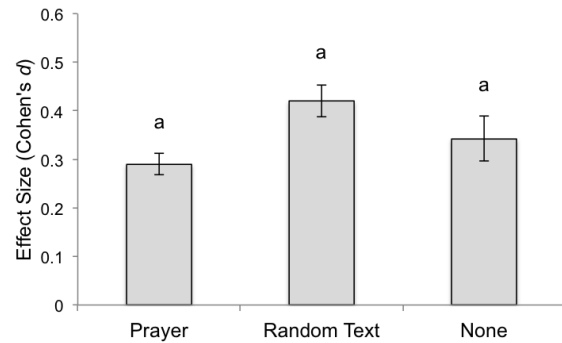
For the religiously-oriented scientists, your prayers for significance can be answered, but you have to work for it. Your prayers will be more successful before you gather your data, which means you still have to do the careful work of proper experimental design, performing the actual experiment, and conducting the appropriate statistical methods. This finding parallels the advice of many scientists that proper planning of an experiment is tantamount to its success (e.g., Festing et al. 2002, BMJ, Montgomery 2013). Just don't expect to use prayer as a last resort when confronted with a messy dataset with which you want to do some data dredging. God does not condone fishing.

In a way, this outcome may have been the best possible result. It's a good thing that scientists (or anybody) can't just pray for a statistically significant result after data is collected. That ability would more or less render the entire Scientific Method meaningless. A common response to "Why doesn't God answer my prayers?" is that "God created us and knows infinitely more than we know. He knows what is best for us, and what would not be good for us." (source) And what wouldn't be good for us is the complete dismantling of the Scientific Method. Either that, or God is really against null hypothesis testing.